Diffusion Models in Vision: A Survey

Abstract

Denoising diffusion models represent a recent emerging topic in computer vision, demonstrating remarkable results in the area of generative modeling. A diffusion model is a deep generative model that is based on two stages, a forward diffusion stage and a reverse diffusion stage. In the forward diffusion stage, the input data is gradually perturbed over several steps by adding Gaussian noise. In the reverse stage, a model is tasked at recovering the original input data by learning to gradually reverse the diffusion process, step by step. Diffusion models are widely appreciated for the quality and diversity of the generated samples, despite their known computational burdens, i.e., low speeds due to the high number of steps involved during sampling. In this survey, we provide a comprehensive review of articles on denoising diffusion models applied in vision, comprising both theoretical and practical contributions in the field. First, we identify and present three generic diffusion modeling frameworks, which are based on denoising diffusion probabilistic models, noise conditioned score networks, and stochastic differential equations. We further discuss the relations between diffusion models and other deep generative models, including variational auto-encoders, generative adversarial networks, energy-based models, autoregressive models and normalizing flows. Then, we introduce a multi-perspective categorization of diffusion models applied in computer vision. Finally, we illustrate the current limitations of diffusion models and envision some interesting directions for future research.